UNCLASSIFIED

# AD 257 503

*Reproduced by the*

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

THE UNIVERSITY OF WISCONSIN

*madison, wisconsin*

UNITED STATES ARMY

# MATHEMATICS RESEARCH CENTER

MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY
THE UNIVERSITY OF WISCONSIN

Contract No.: DA-11-022-ORD-2059

# ON THE USE OF A DISTRIBUTION-FREE PROPERTY IN DETERMINING A TRANSFORMATION OF ONE VARIATE SUCH THAT IT WILL EXCEED ANOTHER WITH A GIVEN PROBABILITY

Sam C. Saunders

MRC Technical Summary Report #225
March 1961

Madison, Wisconsin

# ON THE USE OF A DISTRIBUTION-FREE PROPERTY IN DETERMINING A TRANSFORMATION OF ONE VARIATE SUCH THAT IT WILL EXCEED ANOTHER WITH A GIVEN PROBABILITY

Sam C. Saunders


## 0. Summary

Let $X$ and $Y$ be independent random variables, with continuous distributions $F \in \mathfrak{F}$ and $G \in \mathfrak{G}$, on the sample space $\mathfrak{X}$. Let $\omega$ be a homeomorphism from $\mathfrak{X}$ onto itself and define

$$H(\omega) = \int F(\omega)dG \ .$$

From samples of $X$ and $Y$ we form $\tilde{F}$ and $\tilde{G}$ which are estimates of $F$ and $G$, respectively, and define an estimate of $H$, say $\tilde{H}$,

$$\tilde{H}(\omega) = \int \tilde{F}(\omega)d\tilde{G}$$

for $\omega \in \Omega$, a class of homeomorphisms linearly ordered by $H$.

Interpreting $\omega(Y)$ as the strain under use $\omega$ and $X$ as the strength (of some device), then $H(\omega) = P[X < \omega(Y)]$ represents the unreliability and $\tilde{H}(\omega)$ is an estimate of it. If we use $\tilde{H}$ to determine a use $\tilde{\omega}$, then what is the probability that the true unreliability $H(\omega)$ is too large? We examine this problem under the assumptions that $\tilde{F}(F^{-1})$ and $\tilde{G}(G^{-1})$ are distribution-free with respect to $\mathfrak{F}$ and $\mathfrak{G}$ respectively. This provides an answer in some cases and allows one to obtain stochastic bounds in others.


---

Sponsored by the Mathematics Research Center, United States Army under Contract No.: DA-11-022-ORD-2059.

## 1. Genesis of the Problem

We were asked to consider the problem of determining the unreliability of a missile fuel tank for a given weight of liquid hydrocarbon fuel to be installed. The volume of the tank, the specific weight of a batch of fuel and the temperature environment were stochastic variables.

In the temperature-pressure range for this specific problem considerations of the physical and chemical properties of the fuel show that the design pressure is exceeded if the following event obtains: $[X < wg(Y)]$, where $X$ is the random volume of the tank, $w$ is the weight of the fuel installed and $g$ is a known function (determined by the bulk modulus of the fuel, the temperature variation of range and the design pressure of the tank) of $Y$ the random specific weight of the fuel batch.

If a weight $w$ of fuel is installed, the design unreliability of the tank, i.e., the probability of the pressure exceeding the design specification, is $H(w) = P[X < wg(Y)]$. If the distribution of $X$ was $F$ and that of $Y$ was $G$, both known, then we could express

$$H(w) = \int_0^{\infty} F(wg(x))dG(x) .$$

This equation defines the function $H$ and hence

(a) for a given weight $w$ of fuel installed we can determine the design unreliability $H(w)$,

(b) for a specified design unreliability of at most $\epsilon$ we can seek the maximum weight $w$ for which we have $H(w) \leq \epsilon$.

However, in practice the distributions F and G are usually not known and we have only sample values of tank volumes and the sample values of the specific weights of fuel batches with which to arrive somehow at answers which correspond to the cases (a) and (b).

It is the study of the statistical problems for the situations (a) and (b), within the more inclusive formulation of the problem that we propose, which constitutes the subject matter of this note.

## 2. Introduction and Related Results

Let $(\mathfrak{X}, \mathfrak{A})$ be a measurable space for which $(\mathfrak{X}, \prec)$ is a partially ordered set, i.e., the relation is reflexive, transitive and such that $x \prec y$ and $y \prec x$ imply $x = y$. If the relation is measurable, i.e., for all $x \in \mathfrak{X}$ the set $\{y \in \mathfrak{X} : y \prec x\}$ is in $\mathfrak{A}$, then a distribution can be defined on $\mathfrak{X}$ for each measure P on $\mathfrak{A}$ by $F(x) = P[X \prec x]$, with the usual interpretation of X as the identity random variable (r.v.).

If for each $x \in \mathfrak{X}$ the set $[X = x]$ has measure zero, then the distribution F is continuous. Now $\mathfrak{X}$ is called a positive sample space if and only if (iff) for each $x \prec y$, $x \neq y$ we have the set $[X \succ x] \cap [X \prec y]$ has positive P-measure.

Let X and Y be random variables taking values in the positive sample space $\mathfrak{X}$ with continuous distributions F and G, respectively. (The generalization to Y taking values in a space different from $\mathfrak{X}$ will be seen to be immediate.) Let $\Omega$ be a set of transformations on $\mathfrak{X}$ onto itself. For each $\omega \in \Omega$ we have the probability that $X$ precedes $\omega(Y)$ according to given by

(2.1) $$H(\omega) = P[X \prec \omega(Y)] = \int F\omega \, dG$$

where we make the convention that juxtaposition of functions refers to composition and integrals are understood to be over the entire space $\mathfrak{X}$.

Let us assume that while $F$ and $G$ are unknown we can obtain samples of $X$ and $Y$ from which we form, respectively, the estimates $\widetilde{F}$ and $\widetilde{G}$ of the distributions. Thus we define an estimate of $H$, say $\widetilde{H}$, by

(2.2) $$\widetilde{H}(\omega) = \int \widetilde{F}\omega \, d\widetilde{G}$$

for each $\omega \in \Omega$ for which the integrals exist.

Some questions of interest are: In what sense is $\widetilde{H}$ a good estimate of $H$? If $\omega$ is known, what is the distribution of $\widetilde{H}(\omega)$? But primarily we want the maximum transformation $\widetilde{\omega}$ (as a function of $\widetilde{H}$) for which for each $F \in \mathfrak{F}$, $G \in \mathfrak{G}$ we have

$$P[H(\widetilde{\omega}) > \epsilon] \leq \alpha$$

where $\alpha$ and $\epsilon$ are small and specified in advance. That is, we are interested in the problem of using $\widetilde{H}$ to obtain what would be tolerance limits were $H$ a probability distribution on $\Omega$.

If $\Omega$ consists of a single point $\omega$, we can without loss of generality assume it is the identity transformation. We then have the problem of estimating $p = P[X \prec Y]$ from samples of $X$ and $Y$.

Suppose that $\mathfrak{X}$ is the real line with the usual ordering and $\mathfrak{F} = \mathfrak{G}$

is the class of continuous distributions. Then it is well known that the empiric distribution, say $\hat{G}$, is the unique minimum variance unbiased estimate of $G$. But, further, $\hat{p}$ defined analogously to (2.2) is the unique minimum variance unbiased estimate of $p$ and $mn\hat{p}$ is the Mann-Whitney statistic where $m$ and $n$ are the sample sizes used in computing $\hat{F}$ and $\hat{G}$.

The problem analogous to ours, i.e., involving $\hat{p}$ and $p$, has been studied and bounds for the corresponding estimates have been obtained for large sample sizes under the assumption that one of the distributions is known (Birnbaum [1]) and assuming that neither of the distributions is known (Birnbaum and McCarty [2]).

## 3. Results

Let $\leq$ be a partial ordering on the set of all transformations of onto itself for which $(\Omega, \leq)$ is a linearly ordered subset which is order complete. That is, $(\Omega, \leq)$ is a partially ordered set for which any two elements are comparable and each non-void subset of $\Omega$ which has a lower bound has an infimum.

For a sample of independent. r.v.'s each with the same distribution $F \epsilon \mathfrak{F}$, the estimate $\tilde{F}$ (which is a measurable function of the sample) of $F$ is <u>ample for $\mathfrak{F}$</u> iff the random function $\tilde{F}F^{-1}$ has the same probability law for every $F \epsilon \mathfrak{F}$.

We now make our assumptions:

1° $\mathfrak{F}$ and $\mathfrak{G}$ are classes of continuous distributions on the partially ordered positive sample space $(\mathfrak{X}, \prec)$.

2° $(\Omega, \leq)$ is a linearly ordered complete space of transformations on $\mathfrak{X}$ such that

(3.1) $$\omega_1 \leq \omega_2 \text{ implies } \omega_1(x) \prec \omega_2(x) \text{ for all } x \in \mathcal{X} .$$

3° $\widetilde{F}$ and $\widetilde{G}$ are ample estimates for $F \in \mathcal{F}$ and $G \in \mathcal{G}$, respectively.

Since there always exists a set of transformations on $\mathcal{X}$, say $\Gamma$, such that $\mathcal{F} = \{F_0\gamma : \gamma \in \Gamma\}$ and similarly $\mathcal{G} = \{G_0\lambda : \lambda \in \Lambda\}$ where $\Gamma$ and $\Lambda$ each contain the identity function, it is seen in this representation that $\widetilde{F}$ is ample iff $\widetilde{F}\gamma^{-1}$ (or equivalently iff $\widetilde{\gamma}\gamma^{-1}$) is distribution-free with respect to $\mathcal{F}$. We shall adopt this representation wherever convenient, without comment, in what follows.

The set

(3.2) $$\Phi = \{\gamma\omega\lambda^{-1} : \lambda \in \Lambda , \gamma \in \Gamma , \omega \in \Omega\}$$

is a set of transformations on $\mathcal{X}$ which is, in general, partially ordered by $\leq$. Let

(3.3) $$H_0(\phi) = \int F_0\phi \, dG_0 ,$$

(3.4) $$\widetilde{H}_0(\phi) = \int \widetilde{F}\gamma^{-1}\phi \, d\widetilde{G}\lambda^{-1}$$

be two functions, the second random, defined for each $\phi \in \Phi$ for which the integrals exist. (The functions $\widetilde{F}$ and $\widetilde{G}$ may both contain discontinuities and these cannot be made to coincide.)

We now define for $\delta$ an element of the range of $\widetilde{H}$

(3.5) $$\widetilde{\omega}_\delta = \inf \{\omega \in \Omega : \widetilde{H}(\omega) \geq \delta\}$$

which by completeness of $\Omega$ is a r.v. Let $\epsilon$ be in the range of $H_0$ and $H$. We now pick the unique

$$\theta_\epsilon \in \Omega \quad \text{such that } H_0(\theta_\epsilon) = \epsilon \tag{3.6}$$

and there exists a unique

$$\omega_\epsilon \in \Omega \text{ such that } H(\omega_\epsilon) = \epsilon . \tag{3.7}$$

We now have

Theorem 1: Let $1°$, $2°$, $3°$ be true and $H$ and $\widetilde{H}_0$ be defined as in (2.1) and (3.4) respectively. Then if

$$\Omega \supset \Phi \tag{3.7.1}$$

we have

$$P[H(\widetilde{\omega}_\delta) \geq \epsilon] \leq P[\delta \geq \widetilde{H}_0(\theta_\epsilon)] \tag{3.8}$$

with the right hand side of (3.8) being a distribution-free bound for every $(F, G) \in \mathfrak{F} \times \mathfrak{G}$ .

Proof. It is clear by (3.1) with probability one that $\widetilde{H}$ is monotone increasing on $\Omega = \Phi$ (always $\Phi \supset \Omega$). We have by the positivity of the sample space that $H$ is strictly monotone on $\Omega$ and thus we have

$$[H(\widetilde{\omega}_\delta) \geq \epsilon] = [\widetilde{\omega}_\delta \geq \omega_\epsilon] \subsetneq [\delta \geq \widetilde{H}(\omega_\epsilon)] \tag{3.8.1}$$

because $\widetilde{H}(\widetilde{\omega}_\delta) \leq \delta$ . Now since $\gamma\omega_\epsilon\lambda^{-1} = \theta_\epsilon$, by (3.7.1) we have $\widetilde{H}_0(\theta_\epsilon) = \widetilde{H}(\omega_\epsilon)$ that $\widetilde{H}_0(\theta_\epsilon)$ has a distribution independent of $\mathfrak{F} \times \mathfrak{G}$ is clear by $3°$.

Remark: If $\widetilde{H}$ is continuous a.s. then we obtain equality between the probabilities of equation (3.8).

Remark: An example of a situation when the assumption $\Omega \supset \Phi$ is satisfied occurs when $\Omega$ is a semi-group, with respect to composition, of homeomorphisms on $\mathfrak{X}$ and $\Gamma$ and $\Lambda$ are subsets of $\Omega$. We exhibit a trivial instance of this kind in the next section.

We now state

Theorem 2: Let 1°, 2°, 3° be true and $H$ and $H_0$ be defined as in (2.1) and (3.4) respectively. Then if

(3.9) for each $(\gamma, \lambda) \in \Gamma \times \Lambda$ we have $\omega \geq \gamma\omega\lambda^{-1}$ for all $\omega \in \Omega$

then we have

(3.10) $$P[H(\widetilde{\omega}_\delta) \geq \epsilon] \leq P[\delta \geq \widetilde{H}_0(\theta_\epsilon)]$$

with the right-hand side being a distribution-free bound for every $(F, G) \in \mathfrak{F} \times \mathfrak{G}$.

Proof. It is clear by (3.1) that both $H$ and $\widetilde{H}$ are monotone on $\Omega$. Thus we have

$$[H(\widetilde{\omega}_\delta) \geq \epsilon] = [\widetilde{\omega}_\delta \geq \omega_\epsilon] \subset [\delta \geq \widetilde{H}(\omega_\epsilon)] .$$

Now by (3.9) we have $\gamma\omega_\epsilon\lambda^{-1} \leq \omega_\epsilon$ and hence

$$\epsilon = H(\omega_\epsilon) = H_0(\gamma\omega_\epsilon\lambda^{-1}) \leq H_0(\omega_\epsilon)$$

but $\epsilon = H_0(\theta_\epsilon) \leq H_0(\omega_\epsilon)$ and since $\theta_\epsilon$, $\omega_\epsilon \in \Omega$ we have $\theta_\epsilon \leq \omega_\epsilon$ but then a.s. $\widetilde{H}(\omega_\epsilon) \geq \widetilde{H}(\theta_\epsilon)$ and it follows that

$$[\delta \geq \widetilde{H}(\omega_\epsilon)] \subset [\delta \geq \widetilde{H}(\theta_\epsilon)]$$

and the theorem is proved.

Remark: A useful condition which implies assumption (3.9) of theorem 2 above, that $\omega\lambda \geq \gamma\omega$, is

$$\gamma \leq \lambda \quad \text{and either } \lambda\omega \leq \omega\lambda \text{ or } \gamma\omega \leq \omega\gamma . \tag{3.10.1}$$

In most instances this means only that one r.v. is stochastically smaller than the other and that one of the r.v.'s has a distribution which satisfies a convexity type condition.

Let us make the additional assumption

4° $\tilde{\gamma}$, $\tilde{\lambda}$ are 1 - 1 transformations in $\Gamma$ and $\Lambda$ respectively. We can in this case define

$$\hat{H}_0(\phi) = \int F\tilde{\gamma}^{-1}\phi \, dG\tilde{\lambda}^{-1} . \tag{3.11}$$

We remark that by 3° being satisfied, $\hat{H}_0(\phi)$ has a distribution which depends only on $\phi \in \Phi$. Let

$$\Phi_\alpha = \{\phi \in \Phi : P[\hat{H}_0(\phi) \geq \epsilon] = \alpha\} \tag{3.12}$$

For $\alpha \in (0, 1)$, $\Phi_\alpha$ is not empty whenever there exists $\theta_\alpha \in \Omega$ such that

$$P[\hat{H}_0(\theta_\alpha) \geq \epsilon] = \alpha$$

We now state without proof:

Theorem 3: Let 1°, 2°, 3°, 4°, be true with $H$ and $\hat{H}_0$ as defined as in (2.1) and (3.11) respectively then if

(3.13) for each $\phi \in \Phi$ there exists a unique largest element in $\Omega$, say $\phi_\Omega$, such that $\phi_\Omega \leq \phi$.

Then writing $\tilde{\omega} = (\tilde{\gamma}^{-1}\phi\tilde{\lambda})_\Omega$ for $\phi \epsilon \Phi_a$

(3.14) $$P[H(\tilde{\omega}) \geq \epsilon] \leq P[\hat{H}_0(\phi) \geq \epsilon] = a$$

and the right hand side of (3.14) is distribution-free.

Now we define

(3.15) $$\tilde{\omega}_\delta = \sup_{\phi \epsilon \Phi_\delta} (\tilde{\gamma}^{-1}\phi\tilde{\lambda})\Omega$$

as the element to be used in our estimate.

Remark: In case there exist a $\phi_a$ such that $\tilde{\gamma}^{-1}\phi_a\tilde{\lambda} = \tilde{\omega}_a \epsilon \Omega$ we obtain equality between the probabilities in equation (3.14).

4. Examples

Let us first give an application of theorem 1 and take $\mathfrak{X} = (0,\infty)$, $\Omega = \{\omega : \omega(x) = \omega x, \omega > 0\}$, i.e., here we take the transformations $\omega$ to be scalar multiplication by a positive constant to which we give the same designation. Let the orderings $\leq$, $\prec$ on $\Omega$ and $\mathfrak{X}$ be the same and be the usual ordering on the real line. Let $F_0$ and $G_0$ be distributions on $\mathfrak{X}$ and take $\mathfrak{G} = \{G_0\omega : \omega \epsilon \Omega\}$, $\mathfrak{F} = \{F_0\omega : \omega \epsilon \Omega\}$. Without loss of generality we may assume that $EX = 1/\gamma$ and $EY = 1/\lambda$. Now we define $\tilde{F}(x) = F_0(x/\bar{X})$, $\tilde{G}(x) = G_0(x/\bar{Y})$ and these two estimates are ample for $\mathfrak{F}$ and $\mathfrak{G}$, respectively.

The assumptions of theorem one are seen to be satisfied and $\theta$ is the unique element of $\Omega$ such that

(4.1) $$\int_0^\infty F_0(\theta x)\, dG_0(x) = \epsilon$$

but then $\theta_\epsilon = \gamma\omega_\epsilon\lambda^{-1}$ so a.s. we have by the corollary

(4.2) $$\widetilde{H}(\omega_\epsilon) = \widetilde{H}_0(\theta_\epsilon) = \int_0^\infty F_0\left(\frac{\theta_\epsilon x}{\gamma \bar{X}}\right) dG_0\left(\frac{x}{\lambda \bar{Y}}\right)$$

which must needs be tabulated at points of interest.

The integrations (4.1) and (4.2) might be difficult to perform and could require numerical integration techniques depending upon $F_0$ and $G_0$. In order to continue, let us make the mathematically convenient assumption that the densities exist and are given by

$$f_0(x) = \frac{x^{(k/2)-1}e^{-x/2}}{\Gamma(k/2)2^{k/2}}, \quad g_0(x) = e^{-x},$$

where $k$ is a known parameter of the chi-square distribution.

Thus from (4.1) we obtain

$$\int_0^\infty F_0(\theta x)e^{-x}dx = \int_0^\infty e^{-y/\theta} f_0(y)dy = (1 + 2/\theta)^{-k/2}$$

and $\theta_\epsilon = 2/(\epsilon^{-2/k} - 1)$, but fortunately this integration solves (4.2) and we see

$$\widetilde{H}_0(\theta_\epsilon) = (1 + 2\gamma\bar{X}/\theta\lambda\bar{Y})^{-k/2}$$

from which the probability distribution can be found by a simple transformation, using tables of the F distribution, since $\gamma\bar{X}$ has a chi-square distribution with kn degrees of freedom and $2\lambda\bar{Y}$ has a chi-square distribution with 2m degrees of freedom where n and m are the sample sizes used in computing $\bar{X}$ and $\bar{Y}$, respectively.

## 5. Ample Estimates of Distributions

In this note we restrict ourselves to estimates $\widetilde{F}$ of $F$ which are ample. This property could be of some importance in other applications since in particular $\widetilde{F}$ ample for $\mathfrak{F}$ tells us that the analogue of the $D_n$ statistic, say,

$$\widetilde{D} = \sup_{x \in \mathfrak{X}} |\widetilde{F}(x) - F(x)| = \sup_{y \in (0,1)} |\widetilde{F}F^{-1}(y) - y|$$

is distribution-free with respect to $\mathfrak{F}$.

Ampleness of an estimate of a distribution $F$ is a strong property since, in many instances, it allows us to place confidence contours along the entire distribution function. Suppose $\widetilde{F}$ is ample for $\mathfrak{F}$, a set of distributions on the real line, and $\zeta$ is some distribution function on the unit interval which is in the range of $\widetilde{F}F^{-1}$ and for some constant $\beta$ we have

$$\beta = P[\widetilde{F}F^{-1} \leq \zeta] \geq P[\zeta^{-1}\widetilde{F} \leq F]$$

where $\zeta^{-1}(t) = \inf\{x : \zeta(x) \geq t\}$ then $\zeta^{-1}\widetilde{F}$ provides a lower contour for $F$ with $\beta$ confidence.

We have ampleness within those families of distributions on the real line for which one can estimate percentiles and bounds on those estimates from observations in the region of central tendency such as normal, log-normal, exponential, etc. But we may have ampleness for distributions on spaces of higher dimension, as the following two examples show.

Let $\mathfrak{D}$ be the set of continuous distributions on any partially ordered set $(\mathfrak{X}, \prec)$ as defined in section 2. For $y \in \mathfrak{X}$ define $c(\cdot, y)$ to be the indicator (characteristic) function of the set $[Y \prec y]$. Then

$$\hat{G}(x) = \frac{1}{m} \sum_{j=1}^{m} c(x, Y_j) \tag{5.1}$$

is an ample estimate of the distribution $G \in \mathcal{Y}$ from the sample $(Y_1, \ldots, Y_n)$ of independent observations each with distribution $G$. We shall call $\hat{G}$ the empiric cumulative.

Let $\mathcal{X}$ be Euclidean p-space, with $\prec$ the usual ordering, the elements of which we shall take to be column vectors. Let $X$ be $\mathcal{N}(\mu, \Sigma)$ and let $D$ be the non-singular matrix such that $D \Sigma D' = I$. We shall say here, for brevity, that $D$ diagonalizes $\Sigma$. Then we can write $F(x) = F_0[D(x - \mu)]$ for $x \in \mathcal{X}$, where $F_0$ is the distribution of a $\mathcal{N}(0, I)$ variate.

Let $\hat{\mu}$ and $\hat{\Sigma}$ be the usual maximum likelihood estimates taken from a sample of size $n > p$ and define $\hat{D}$ as the diagonalizing matrix for $\hat{\Sigma}$. It exists a.s. Now if we define

$$\sigma(x) = D(x - \mu) \tag{5.2}$$

and the corresponding definition of $\hat{\sigma}$ then fix $y \in \mathcal{X}$ and set

$$T = \hat{\sigma}\sigma^{-1}(y) = \hat{D}D^{-1}[y - D(\hat{\mu} - \mu)] \tag{5.3}$$

that this has a distribution which is the same for all $\mu$, $\Sigma$ is sufficient for ampleness. But clearly $D(\hat{\mu} - \mu)$ is $\mathcal{N}(0, \frac{1}{n} I)$ so it is sufficient that the vector in square brackets in (5.3) satisfies the condition.

Set $\hat{B} = \hat{D}D^{-1}$. We show that this has a distribution independent of $\mu$, $\Sigma$ also. From (5.2) we have

$$\hat{B} D \hat{\Sigma} D' \hat{B}' = I . \tag{5.4}$$

But $nD\hat{\Sigma}D'$ is distributed as $\sum_{i=1}^{n-1} Z_i Z'_i$ where $Z_i$ is $\mathcal{N}(0, I)$ independent of $Z_j (i \neq j)$ .

Hence, except for zero probability; $\hat{B}$ diagonalizes a random matrix which has a distribution independent of $\mu, \Sigma$ . Note that $T'T$ is, except for a scale factor, Hotelling's $T^2$ with non-centrality factor $y'y/n$ .

## 6. Some Remarks on Properties of the Estimate H

Let us set $\overline{F} = E\widetilde{F}$ with similar meaning for $\overline{G}$ and $\overline{H}$. By applying the Fubini theorem and integrating by parts we have for $\omega$

$$\overline{H}(\omega) = \int \overline{F}\omega \, d\overline{G}$$

and

$$H(\omega) - \overline{H}(\omega) = \int F\omega \, d(G - \overline{G}) + \int (F\omega - \overline{F}\omega) \, d\overline{G} .$$

Remark 6.1: $\widetilde{H}$ is an unbiased estimate of $H$ on $\Omega$, if $\widetilde{F}$ and $\widetilde{G}$ are unbiased estimates of $F$ and $G$, respectively.

The converse is not true without additional assumptions. For the real line, unbiasedness of $\widetilde{H}$ on the entire set of 1-1 order preserving maps of onto itself is sufficient for unbiasedness of $\widetilde{F}$ and $\widetilde{G}$ when one imposes the conditions that $\overline{F}$ and $\overline{G}$ are continuous.

We say that $\widetilde{F}_n$ is consistent for $F$ whenever

$$\sup_{\mathfrak{X}} | \widetilde{F}_n - F | \xrightarrow{P} 0 ,$$

where, as before, the subscript $n$ refers to the sample size used to obtain the estimate. We shall refer to consistency exclusively in this sense.

Now

$$\widetilde{H}(\omega) - H(\omega) = \int (G - \widetilde{G})\, d\widetilde{F}\omega + \int (\widetilde{F}\omega - F\omega)\, dG .$$

Thus it follows that

$$|\widetilde{H}(\omega) - H(\omega)| \leq \sup_{\mathfrak{X}} |\widetilde{F} - F| + \sup_{\mathfrak{X}} |\widetilde{G} - G| .$$

Remark 6.2: If $\widetilde{F}_n$ and $\widetilde{G}_m$ are consistent for $F$ and $G$, respectively, then $\widetilde{H}_{mn}$ is consistent for $H$ uniformly on $\Omega$ as $1/n + 1/m \to 0$ .

Remark 6.3: If $\hat{F}_n$ and $\hat{G}_m$ are the empiric cumulatives, then $H_{mn}$, for each $\omega \epsilon \Omega$, is consistent, unbiased and asymptotically normal with $1/n + 1/m \to 0$.

This follows from the known behavior of U-Statistics.

If we assume $\widetilde{F}$ is consistent and $\hat{G}$ is the empiric cumulative, we should obtain asymptotic normality if $\widetilde{F}$ "stabilizes" rapidly enough. We do not attempt to find the weakest such conditions, but we have

Remark 6.4: If $\hat{G}$ is the empiric cumulative, then the function $H$ defined by

$$H_{mn}(\omega) = \frac{1}{m} \sum_{j=1}^{m} \widetilde{F}_n(\omega Y_j)$$

is a consistent estimator of $H$ and asymptotically normal if $1/n + 1/m \to 0$ in such a way that $\sqrt{m} \sup |\widetilde{F}_n - F| \xrightarrow{P} 0$ .

Proof: Let $Z_m = \frac{1}{\sqrt{m}} \sum_{j=1}^{m} [F(\omega Y_j) - H(\omega)]$ and define $\zeta_{m,n}$ by the equation $\sqrt{m}\,[\widetilde{H}_{mn}(\omega) - H(\omega)] = Z_n + \zeta_{m,n}$ . That $Z_m$ is asymptotically normal is clear, and the result follows if $\zeta_{m,n} \xrightarrow{P} 0$ , which is guaranteed by the hypothesis.

7. An application to Reliability Theory

Let $\mathfrak{X} = (0, \infty)$ and let $Y(\omega)$ be the demand time for a particular equipment in a specified environment during an inspection period of length $\omega$. Let $X$ be the life length of this equipment under continued use in that environment.

We want the maximum $\omega$ such that for given $\epsilon > 0$

$$P[X < Y(\omega)] \leq \epsilon \quad .$$

We particularize by making the assumption that $Y(\omega)$ denotes multiplication and without loss of generality we suppose that a scaling factor has been introduced so that $\omega \epsilon (0, 1)$. Now

$$H(\omega) = P[X < \omega Y] = \int F(\omega y) dG(y)$$

where we assume that $X$ and $Y$ are independent r.v.s. with continuous distributions $F$ and $G$ respectively.

Now if $F$ is a distribution with derivative $f$, then the function $\gamma'$ defined by

$$(7.1) \qquad \gamma'(x) = \frac{f(x)}{1 - F(x)} \quad \text{for} \quad x > 0$$

is called the <u>failure rate</u> (or hazard rate).

A common and intuitively appealing assumption concerning life length distributions is

A° the failure rate of $X$ exists and is non-decreasing.

We also assume

B° $X$ is stochastically larger than $Y$, i.e., $F \leq G$.

But, further, without loss of generality we assume

(7.2) $$\frac{1}{2} > \epsilon > 0 \text{ is given and } P[X < Y] > \epsilon \; .$$

Let

$$\gamma(x) = \int_0^x \gamma'(t)dt \quad \text{for } x > 0 \; .$$

Then if we set

$$F_0(x) = G_0(x) = 1 - \epsilon^{-x} \; , \quad x > 0 \; ,$$

from (7.1) we see that $G(x) = G_0 \, \lambda(x)$ .

It is clear that $B°$ implies $\gamma \leq \lambda$ and we now show that $A°$ implies $\gamma\omega \leq \omega\gamma$ and thus (3.10.1) implies assumption (3.9) of theorem 2 . To see this, note that $\dot{\gamma}$ must be convex and hence $\gamma(\omega x) \leq \omega\gamma(x)$ for $\omega \epsilon (0, 1)$ since $\gamma(0) = 0$. This is precisely our condition.

We now seek the unique $\theta$ such that

(7.3) $$\int_0^\infty F_0(\theta x) \, dG_0(x) = \epsilon \; .$$

This may be easily integrated directly or by comparison with (4.1) seen to be $\theta_\epsilon = \frac{\epsilon}{1 - \epsilon}$ .

We now choose $\hat{F}$, $\hat{G}$ to be the empiric cumulatives as defined in (5.1) (which are ample). Hence we need tabulations of the statistic

$$\hat{H}_0(\theta) = \int \hat{F} F^{-1} \theta \, d\hat{G} G^{-1} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} c(\theta U_j, V_i)$$

where $F(X_i) = V_i$ , $i = 1, \ldots, n$, and $G(Y_j) = U_j$ for $j = 1, \ldots, m$ are all independent r.v.'s uniform on (0, 1).

Consider the events

$$A_k = [V_{(k)} < \theta < V_{(k+1)}] \ , \quad k = 0, \ldots, n \ ,$$

where $V_{(1)}, \ldots, V_{(n)}$ are the ordered observations of $V$ with $V_{(o)} = 0$, $V_{(n+1)} = 1$. These events $A_k$ form a disjunct partition of the sample space.

It is clear that $\hat{H}_0(1) = U_{m,n}$ is the well-known Mann-Whitney statistic based on samples of size $m$ and $n$. Now let $S_{mn}(\theta) = mnH_0(\theta)$, and it follows that

$$S_{mn}(\theta) \,|\, A_k = U_{mk} \quad \text{for} \quad k = 0, 1, \ldots, n \ .$$

With the convention that $U_{mo} = 0$ with probability one, we obtain

$$P[S_{mn}(\theta) \le t] = \sum_{k=0}^{n} P[S_{mn}(\theta) \le t \,|\, A_k] P[A_k]$$

$$= \sum_{k=0}^{n} b(k; n, \theta) P[U_{mk} \le t] \ , \quad t = 0, 1, \ldots, mn \ ,$$

We know that $ES_{mn}(\theta) = \frac{mn}{2}\theta$ and since we have

$$EU^2_{mn} = \frac{nm}{12}(n + m + 1) + (nm/2)^2 \ ,$$

it follows by (7.4) directly that

$$ES^2_{mn}(\theta) = \sum_{k=0}^{n} b(k: n, \theta) EU^2_{mk}$$

$$= \frac{m(m+1)n\theta}{12} + \left(\frac{m}{12} + \frac{m^2}{4}\right)[n(n-1)\theta^2 + n\theta] \ .$$

Hence we obtain the variance of $S_{mn}(\theta)$ as

$$Var(S_{mn}(\theta)) = \frac{mn\theta}{6}\left[2m - \frac{3m\theta}{2} + \frac{(n-1)\theta}{2} + 1\right] \ .$$

The results of the preceding section show that $S_{mn}(\theta)$ has asymptotically a normal distribution with the above mean and variance; however, for small sample sizes and small $\theta$ the normal approximation may not be of sufficient accuracy. In the following paragraph we give a few formulae in the range of $t = mn\delta$ small. This is near the region of interest since $\alpha$ is small.

Let $P_{mn}(t) = P[U_{mn} \leq t]$. Now, from results and recurrence formulae in [3], one can obtain the following:

For $m \geq 1$, $k \geq 0$,

$$P_{mk}(0) = 1/\binom{m+k}{k} \qquad \text{for all } k \geq 0 ,$$

$$P_{mk}(1) = \begin{cases} 1 & k = 0 , \\ 2/\binom{m+k}{k} & k \geq 1 , \end{cases}$$

$$P_{mk}(2) = \begin{cases} 1 & \text{if } m \leq 2 \text{ or } k = 0, 1 , \\ 4/\binom{m+k}{k} & \text{if } m \geq 2 \text{ and } k \geq 2 , \end{cases}$$

$$P_{mk}(3) = \begin{cases} 1 & \text{if } m \leq 3 \text{ or } k = 0, 1 , \\ (4+k)/\binom{m+k}{k} & \text{if } m \geq 3 \text{ and } k \geq 2 , \end{cases}$$

and for greater values of $t$ the numerators are functions of higher powers of $k$.

We have the expression

$$F[S_{mn}(\theta) \leq 0] = \sum_{k=0}^{n} \frac{b(k:\, n, \theta)}{\binom{m+k}{k}} = \frac{E(m+n,\, m,\, \theta)}{\binom{m+n}{n}\, \theta^m} , \tag{7.5}$$

where the notation $E$ is that used in the Harvard Tables of the Binomial Distribution [4]. Through the use of these tables, for moderate values of $m$, $n$ and $\theta$, the probability (7.5) can be calculated and for very small values of $\theta$ the first terms of the expansion given can be used for adequate approximation.

Similarly:

for $n \geq 1$, $m \geq 1$,

$$P[S_{mn}(\theta) \leq 1] = (1-\theta)^n + \frac{2E(m+n,\ m+1,\ \theta)}{\binom{m+n}{m}\theta^m},$$

for $n \geq 1$, $m \geq 2$,

$$P[S_{mn}(\theta) \leq 2] = (1-\theta)^n + n\theta(1-\theta)^{n-1} + \frac{E(m+n,\ m+2,\ \theta)}{\binom{m+n}{m}\theta^m}.$$

We now return to the reliability problem. From the formulae preceding (or others similarly developed) we discover, through tabulation, the $\alpha$th percentile of the r.v. $S_{mn}(\theta)$, thus we find $\delta_\alpha$, by dividing by mn, such that

$$P[\hat{H}_0(\theta) \leq \delta_\alpha] = \alpha .$$

Then we set

$$\hat{\omega}_\delta = \inf\{\omega \epsilon (0,1): \hat{H}(\omega) \geq \delta_\alpha\}$$

which is the empiric derating percentage sought. We know by the assumption A° and B° that the theorem applies and thus

$$P[H(\hat{\omega}_\delta) \geq \epsilon] \leq \alpha .$$

In this case, for example, if we chose $\delta = 0$ we would find that

$$\hat{\omega}_0 = \frac{\min X_i}{\max Y_j}$$

which is intuitively appealing.

# BIBLIOGRAPHY

[1] Z. W. Birnbaum, "On a Use of the Mann-Whitney Statistic", Proc. of Third Berkeley Symposium, Vol. 1 (1956), pp. 13-17.

[2] Z. W. Birnbaum and R. C. McCarty, "A Distribution-free Upper Confidence Bound for $P[Y < X]$, Based on Independent Samples of $X$ and $Y$", Annals of Math. Stat. Vol. 29 (1958), pp. 558-562.

[3] H. B. Mann and D. R. Whitney, "On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other", Annals of Math. Stat., Vol. 18, (1947), pp. 50-60.

[4] Tables of the Cumulative Binomial Probability Distribution, Harvard University Press, 1955.